

550971

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION  
EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété  
Intellectuelle  
Bureau international



(43) Date de la publication internationale  
14 octobre 2004 (14.10.2004)

PCT

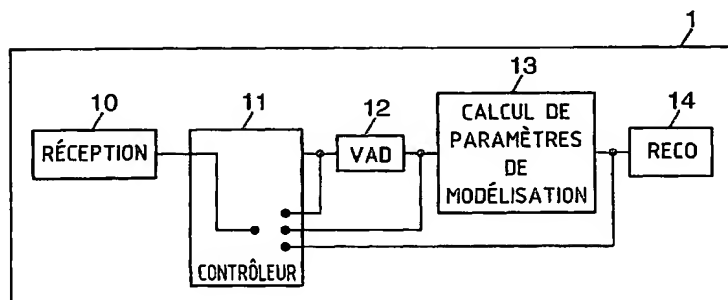
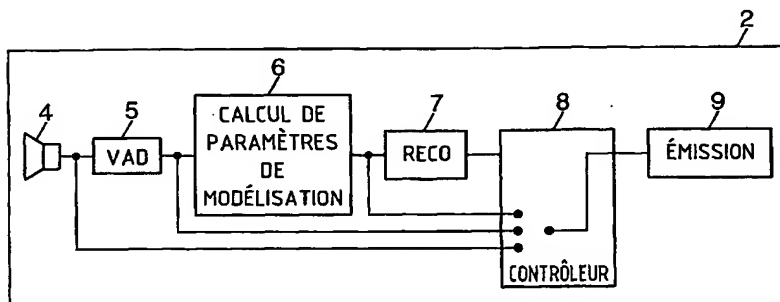
(10) Numéro de publication internationale  
WO 2004/088637 A1

- (51) Classification internationale des brevets<sup>7</sup> : G10L 15/28 (72) Inventeurs; et  
(21) Numéro de la demande internationale : PCT/FR2004/000547 (75) Inventeurs/Déposants (pour US seulement) : MONNE, Jean [FR/FR]; 53, rue du Pré de Saint-Maur, F-22700 Perros Guirec (FR). PETIT, Jean-Pierre [FR/FR]; 10, lot Zant Erwan, F-22220 Minihiy Tréguier (FR). BRISARD, Patrick [FR/FR]; 5, allée du Cèdre, F-92320 Châtillon (FR).  
(22) Date de dépôt international : 8 mars 2004 (08.03.2004)  
(25) Langue de dépôt : français  
(26) Langue de publication : français  
(30) Données relatives à la priorité : 03/03614 25 mars 2003 (25.03.2003) FR  
(71) Déposant (pour tous les États désignés sauf US) : FRANCE TELECOM [FR/FR]; 6, place d'Alleray, F-75015 Paris (FR). (74) Mandataires : LOISEL, Bertrand etc.; Cabinet Plasseraud, 65/67, rue de la Victoire, F-75440 Paris Cedex 09 (FR).  
(81) États désignés (sauf indication contraire, pour tout titre de protection nationale disponible) : AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO,

[Suite sur la page suivante]

(54) Title: DISTRIBUTED SPEECH RECOGNITION METHOD

(54) Titre : PROCÉDE DE RECONNAISSANCE DE PAROLE DISTRIBUEE



6 / 13 CALCULATION OF MODELLING PARAMETERS

8 / 11 CONTROLLER

9 TRANSMISSION

10 RECEPTION

(57) Abstract: The invention relates to a distributed speech recognition method comprising at least one user terminal and at least one server which can communicate with each other by means of a telecommunication network. The inventive method comprises the following steps consisting in: at the user terminal, attempting to associate a saved form with the signal to be recognised and, independently of said step, transmitting a signal to the server, indicating the signal to be recognised; and, at the server, attempting to associate a saved form with the signal received.

(57) Abrégé : Procédé de reconnaissance de parole distribuée, comportant au moins un terminal utilisateur et au moins un serveur aptes à communiquer entre eux par l'intermédiaire d'un réseau de télécommunications, selon lequel au niveau du terminal utilisateur, on tente d'associer une forme mémorisée au signal à reconnaître ; et indépendamment de l'étape de tentative d'association d'une forme mémorisée, on émet un signal à destination du serveur indiquant le signal à reconnaître. Au niveau du

serveur, on tente d'associer une forme mémorisée au signal reçu.

WO 2004/088637 A1



CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

- (84) États désignés (sauf indication contraire, pour tout titre de protection régionale disponible) : ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), eurasién (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Déclaration en vertu de la règle 4.17 :**

- relative à la qualité d'inventeur (règle 4.17.iv)) pour US seulement

**Publiée :**

- avec rapport de recherche internationale  
— avant l'expiration du délai prévu pour la modification des revendications, sera republiée si des modifications sont reçues

En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.

## PROCEDE DE RECONNAISSANCE DE PAROLE DISTRIBUEE

La présente invention est relative au domaine de la commande vocale d'applications, exercée sur des terminaux utilisateurs, grâce à la mise en oeuvre de moyens de reconnaissance de la parole. Les terminaux utilisateurs considérés sont tous les dispositifs dotés d'un moyen de capture de la parole, communément un microphone, possédant des capacités de traitement de ce son et reliés à un ou des serveurs par un canal de transmission. Il s'agit par exemple d'appareils de commande, de télécommande utilisés dans des applications domotiques, dans des automobiles (commande d'auto-radio ou d'autres fonctions du véhicule), dans des PC ou des postes téléphoniques. Le champ des applications concernées est essentiellement celui où l'utilisateur commande une action, demande une information ou veut interagir à distance en utilisant une commande vocale. L'utilisation de commandes vocales n'exclut pas l'existence dans le terminal utilisateur d'autres moyens d'action (système multi-modal), et le retour d'informations, d'états ou de réponses peut également se faire sous forme combinée visuelle, sonore, olfactive et tout autre moyen humainement perceptif.

De manière générale, les moyens pour la réalisation de la reconnaissance de parole comprennent des moyens d'obtention d'un signal audio, des moyens d'analyse acoustique qui extraient des paramètres de modélisation et enfin des moyens de reconnaissance qui comparent ces paramètres de modélisation extraits à des modèles, et proposent la forme mémorisée dans les modèles qui peut être associée au signal de la façon la plus probable. Optionnellement des moyens de détection d'activité vocale VAD ("Voice Activation Detection") peuvent être utilisés. Ils assurent la détection des séquences correspondant à de la parole et devant être reconnues. Ils extraient du signal audio en entrée, en-dehors des périodes d'inactivité vocale, des segments de parole, qui seront ensuite traités par les moyens d'extraction de paramètres de modélisation.

Plus particulièrement, l'invention porte sur les interactions entre les trois modes de reconnaissance de la parole dits embarqué, centralisé et distribué.

Dans un mode de reconnaissance de parole embarquée, l'ensemble  
5 des moyens pour effectuer la reconnaissance de parole se trouvent au niveau du terminal utilisateur. Les limitations de ce mode de reconnaissance sont donc liées notamment à la puissance des processeurs embarqués, et à la mémoire disponible pour stocker les modèles de reconnaissance de parole. En  
10 contrepartie, ce mode autorise un fonctionnement autonome, sans connexion à un serveur, et à ce titre est voué à un fort développement lié à la réduction du coût de la capacité de traitement.

Dans un mode de reconnaissance de la parole centralisée, toute la  
procédure de reconnaissance de parole et les modèles de reconnaissance se trouvent et s'exécutent sur une machine, appelée généralement serveur vocal,  
15 accessible par le terminal utilisateur. Le terminal transmet simplement au serveur un signal de parole. Cette méthode est utilisée notamment dans les applications offertes par les opérateurs de télécommunication. Un terminal basique peut ainsi accéder à des services évolués, activés à la voix. De nombreux types de reconnaissance de parole (robuste, flexible, très grand  
20 vocabulaire, vocabulaire dynamique, parole continue, mono ou multi locuteurs, plusieurs langues, etc ) peuvent être implémentés dans un serveur de reconnaissance de parole. En effet, les machines centralisées ont des capacités de stockage de modèles, des tailles de mémoire de travail et des puissances de calcul importantes et croissantes.

25 Dans un mode de reconnaissance de parole distribuée, les moyens d'analyse acoustique sont embarqués dans le terminal utilisateur, les moyens de reconnaissance étant au niveau du serveur. Dans ce mode distribué, une fonction de débruitage associée aux moyens d'extraction des paramètres de modélisation peut être avantageusement réalisée à la source. Seuls les  
30 paramètres de modélisation sont transmis, ce qui permet un gain substantiel en débit de transmission, particulièrement intéressant pour les applications multimodales. De plus, le signal à reconnaître peut être mieux protégé contre les erreurs de transmission. Optionnellement on peut aussi embarquer la

détection d'activité vocale (VAD) pour ne transmettre les paramètres de modélisation que durant les séquences de parole, ce qui a pour avantage de réduire de manière importante la durée de transmission active. La reconnaissance de parole distribuée permet en outre de véhiculer sur le même

5 canal de transmission des signaux de parole et de données, notamment texte, images ou vidéos. Le réseau de transmission peut être par exemple de type IP, GPRS, WLAN ou Ethernet. Ce mode permet également de bénéficier de procédures de protection et de correction contre les pertes de paquets constituant le signal transmis à destination du serveur. Cependant il nécessite

10 la disponibilité de canaux de transmission de données, avec un protocole strict de transmission.

L'invention propose un système de reconnaissance de parole comportant des terminaux utilisateurs et des serveurs combinant les différentes fonctions offertes par les modes de reconnaissance de parole embarquée,

15 centralisée et distribuée, pour offrir le maximum d'efficacité, de confort et d'ergonomie aux utilisateurs de services multi modaux où la commande vocale est utilisée.

Le brevet US 6 487 534-B1 décrit un système de reconnaissance de parole distribuée comportant un terminal utilisateur disposant des moyens de

20 détection d'activité vocale, de moyens d'extraction des paramètres de modélisation et de moyens de reconnaissance. Ce système comprend en outre un serveur disposant également de moyens de reconnaissance. Le procédé décrit est consistant à effectuer une première phase de reconnaissance au niveau du terminal utilisateur. Puis en fonction des résultats de cette première

25 phase, les paramètres de modélisation calculés au niveau du terminal peuvent être alors envoyés à destination du serveur, afin notamment de déterminer cette fois grâce aux moyens de reconnaissance du serveur, une forme mémorisée dans les modèles de celui-ci et associée au signal envoyé.

Le but visé par le système décrit dans le document cité est de diminuer

30 la charge au niveau du serveur. Cependant il s'ensuit que la reconnaissance s'effectue au mieux après le temps nécessaire à la première phase dans le terminal. Lorsqu'une deuxième phase doit avoir lieu, le temps de reconnaissance total est égal au temps de reconnaissance de la première

phase ajouté à celui de la seconde phase.

Un but de la présente invention est de proposer un procédé qui vise à optimiser le temps nécessaire à la reconnaissance de parole, et donc à la réalisation de l'action demandée par l'utilisateur.

5        Suivant un premier aspect, l'invention propose un procédé de reconnaissance de parole distribuée, comportant au moins un terminal utilisateur et au moins un serveur, aptes à communiquer entre eux par l'intermédiaire d'un réseau de télécommunications, selon lequel au niveau du terminal utilisateur, on réalise au moins les étapes suivantes :

10                - obtenir un signal audio à reconnaître ; et,  
                  - calculer des paramètres de modélisation du signal audio à reconnaître; et,  
                  - tenter d'associer une forme mémorisée aux paramètres de modélisation; et,

15                - indépendamment de l'étape de tentative d'association d'une forme mémorisée, émettre à destination du serveur un signal indiquant le signal audio à reconnaître;

et au niveau du serveur, on réalise au moins les étapes suivantes :

                  - recevoir le signal émis par le terminal utilisateur ;  
20                - tenter d'associer une forme mémorisée au signal reçu.

Un procédé selon l'invention permet d'éviter de cumuler un temps de traitement au niveau du terminal et un temps de traitement au niveau du serveur, et ainsi de diminuer le temps pour donner suite à la demande de l'utilisateur.

25        Il permet aussi de tirer avantage des avantages des caractéristiques des moyens de reconnaissance dont disposent respectivement le terminal et le serveur, notamment lorsque le signal à reconnaître est de nature indéfinie pour déterminer au plus vite la réponse à effectuer.

30        Dans des modes de mise en oeuvre préférés, le signal émis par le terminal utilisateur à destination du serveur est sélectionné parmi au moins le signal audio à reconnaître et un signal indiquant les paramètres de

modélisation. Alors, si le signal reçu est de type audio, le serveur calcule des paramètres de modélisation du signal audio reçu; et tente d'associer une forme mémorisée aux paramètres de modélisation du signal audio reçu. Si le signal reçu indique des paramètres de modélisation, le serveur tente d'associer une  
5 forme mémorisée auxdits paramètres de modélisation.

Le choix du signal transmis - soit le signal audio (compressé ou non), soit le signal délivré par les moyens de calcul des paramètres de modélisation du terminal - peut être défini soit par le type d'applications en cours, soit par l'état du réseau, soit suite à une coordination entre des moyens de contrôle  
10 respectifs du terminal et du serveur.

Avantageusement, l'obtention au terminal du signal à reconnaître comporte une détection d'activité vocale appliquée à un signal audio d'origine pour produire le signal audio à reconnaître en débarrassant le signal d'origine de périodes d'inactivité vocale. Le signal à émettre sera donc sélectionné  
15 parmi ce signal audio débarrassé de périodes d'inactivité vocale et le signal indiquant des paramètres de modélisation.

Dans un mode de mise en œuvre du procédé selon l'invention, le signal émis est sélectionné un signal parmi au moins le signal audio d'origine, le signal audio représentatif du signal d'origine débarrassé de périodes d'inactivité vocale après détection vocale et le signal indiquant les paramètres de modélisation.  
20

Avantageusement, si le signal reçu est un signal audio débarrassé de périodes d'inactivité vocale, le serveur calcule des paramètres de modélisation du signal reçu et tente d'associer une forme mémorisée aux paramètres de modélisation du signal audio reçu. Lorsque le signal reçu est un signal de type audio, mais sur lequel il n'a pas été réalisée de détection d'activité vocale, le serveur réalise une détection d'activité vocale appliquée au signal audio reçu pour produire un signal audio à reconnaître en débarrassant le signal d'origine de périodes d'inactivité vocale. Puis il calcule des paramètres de modélisation du signal audio à reconnaître. Enfin, il tente d'associer une forme mémorisée aux paramètres de modélisation.  
25  
30

Avantageusement, on choisit la forme mémorisée associée déterminée au niveau du terminal quand elle existe. On choisit la forme mémorisée

associée déterminée la première. Ou encore on choisit la forme mémorisée associée la meilleure selon un critère défini (par exemple, probabilité de vraisemblance).

5        Suivant un deuxième aspect, l'invention propose un terminal utilisateur pour mettre en œuvre le procédé de reconnaissance de parole distribuée décrit ci-dessus.

      Suivant un troisième aspect, l'invention propose un serveur pour mettre en œuvre le procédé de reconnaissance de parole distribuée décrit ci-dessus.

10       Dans un mode de réalisation préféré, certains au moins des moyens pour réaliser le traitement de la reconnaissance au niveau du terminal (moyens d'extraction de paramètres ou des moyens de reconnaissance) ont été téléchargés par l'intermédiaire du réseau de télécommunications. Ils peuvent par exemple être téléchargés par le serveur.

15       D'autres caractéristiques et avantages de l'invention apparaîtront encore à la lecture de la description qui va suivre. Celle-ci est purement illustrative et doit être lue en regard des dessins annexés sur lesquels la figure unique est un schéma représentant un exemple de terminal utilisateur et un exemple de serveur selon l'invention.

20       Le système représenté sur la figure unique comporte un serveur 1 et un terminal utilisateur 2, qui communiquent entre eux par l'intermédiaire d'un réseau (non représenté) disposant de canaux pour la transmission de signaux de voix et de canaux pour la transmission de signaux de données.

      Le terminal 2 comporte un microphone 4, qui recueille la parole à reconnaître d'un utilisateur sous forme d'un signal audio. Le terminal 2  
25       comporte également des modules 5, 6, 7. Le module de détection d'activité vocale VAD 5 assure la détection des séquences correspondant à de la parole et devant être reconnues. Ce module 5 est par exemple conçu pour détecter rapidement des mots de commandes. Le module 6 effectue une analyse acoustique de façon connue en soi : il calcule des paramètres de modélisation,  
30       tout en réalisant une fonction de débruitage.

      Le module 7 exécute un algorithme de reconnaissance de type connu, par exemple à base de modèles de Markov cachés avec un vocabulaire de



taille réduite. Ce moteur de reconnaissance 7 peut fonctionner en mode monolocuteur, et nécessite une phase d'apprentissage préalable à partir de la voix de l'utilisateur.

Le terminal comprend un contrôleur 8 adapté pour sélectionner, un  
5 signal audio parmi le signal audio en sortie du microphone 4, un signal représentatif des segments de parole extraits par les moyens de détection d'activité vocale 5 et un signal indicatif des paramètres de modélisation 6.

Le terminal comprend en outre une interface 9 pour l'émission sur le réseau, à destination du serveur, du signal sélectionné par le contrôleur 8.

10 Le serveur 1 comporte une interface réseau 10 pour recevoir les signaux qui lui sont adressés et un contrôleur 11 qui analyse le signal reçu et le dirige ensuite sélectivement vers un module de traitement du serveur parmi plusieurs modules 12,13,14. Le module 12 est un détecteur d'activité vocale, qui assure la détection des segments correspondant à de la parole de façon  
15 similaire au module 5. Toutefois, il peut être différent du module 5, et par exemple conçu pour détecter rapidement des phrases entières. Son temps de réaction peut donc être différent de celui du module 5. Dans l'exemple présent, son temps de réaction sera plus lent. Le module 13 assure le calcul de paramètres de modélisation de façon semblable au module de calcul 6 du  
20 terminal. Toutefois, le modèle de calcul peut être différent. Le module 14 exécute un algorithme de reconnaissance de type connu, par exemple à base de modèles de Markov cachés avec un vocabulaire de taille quelconque, par exemple supérieur à 100 000 mots. Ce moteur de reconnaissance 14 compare les paramètres en entrée à des modèles de parole qui représentent des mots  
25 ou des phrases, et détermine la meilleure forme associée, compte tenu de modèles syntaxiques qui décrivent les enchaînements de mots attendus, de modèles lexicaux qui précisent les différentes prononciations des mots, et de modèles acoustiques représentatifs des sons prononcés. Ces modèles sont par exemple multilocuteurs, capables de reconnaître, avec une bonne fiabilité,  
30 de la parole, indépendamment du locuteur.

Le contrôleur 11 commande le module de VAD 12, le module de calcul de paramètres 13 et le moteur de reconnaissance 14 de façon à :

a/ lorsque le signal reçu par l'interface réseau 10 est de type audio et n'indique pas de segments de parole obtenus après détection d'activité vocale, activer le module de détection d'activité vocale 12 en leur adressant le signal reçu en tant que signal d'entrée, puis adresser les segments de parole extraits  
5 par le module 12 au module de calcul de paramètres de modélisation 13 en tant que paramètres d'entrée, puis adresser les paramètres extraits par ce module 13 au moteur de reconnaissance 14 en tant que paramètres d'entrée ;

b/ lorsque le signal reçu par l'interface de réception 10 indique des segments de parole après détection d'activité vocale, activer le module de  
10 calcul de paramètres de modélisation 13 en lui adressant le signal reçu en tant que signal d'entrée, puis adresser les paramètres extraits par ce module 13 au moteur de reconnaissance 14 en tant que paramètres d'entrée ;

c/ lorsque le signal reçu par l'interface de réception 10 indique des paramètres de modélisation, adresser lesdits paramètres indiqués au moteur  
15 de reconnaissance 14 en tant que paramètres d'entrée.

Considérons une application dans laquelle l'utilisateur énonce :  
« appelle Antoine », Antoine figurant dans le répertoire local. Le signal audio correspondant obtenu par le microphone 4 du terminal est traité par le module  
20 VAD 5, qui en extrait des segments de parole qui sont à leur tour adressés aux module 6, qui calcule des paramètres de modélisation. Ces paramètres sont ensuite adressés au moteur de reconnaissance 7 pour y associer une forme issue du dictionnaire local. En parallèle, le contrôleur 8 a sélectionné un signal à émettre parmi le signal audio d'origine, un signal audio indiquant les segments de parole extraits du signal audio d'origine après détection d'activité  
25 vocale et un signal indiquant les paramètres de modélisation. Le signal sélectionné est émis à l'aide de l'interface d'émission 9 à destination du serveur.

Dans le mode de réalisation considéré, le signal sélectionné par le contrôleur du terminal 8 est le signal audio d'origine, qui a été envoyé en  
30 direction du serveur dès qu'il a été acquis par le microphone 4.

Le serveur réceptionne le signal à reconnaître envoyé par le terminal, et le traite comme indiqué en a/ ci-dessus.

Le processus de reconnaissance s'effectue ainsi de part et d'autre. Le terminal détermine dans un temps T1 une forme associée F1; le serveur détermine une autre forme associée F2 dans un temps T2 différent de T1. Une des deux formes est retenue selon un critère de choix. Le critère de choix peut  
5 est par exemple le suivant : on choisit la forme la plus rapidement trouvée et ce dès qu'elle est trouvée, sans attendre la détermination de l'autre forme. L'applicatif local au terminal passe alors à la phase applicative suivante.

Puis l'utilisateur énonce « rechercher le message de Josiane ».

La réponse au niveau du terminal n'est pas consistante et se solde par  
10 un rejet après un temps T1'. La reconnaissance s'effectue en parallèle au niveau du serveur et aboutit en un temps T2' à la détermination d'une forme associée, qui va permettre de jouer le message demandé par l'utilisateur.

Un procédé de reconnaissance selon l'invention permet ainsi de combiner les avantages des systèmes de reconnaissance au niveau du  
15 terminal et du serveur. Les mots courts sont rapidement déterminés par le moteur de reconnaissance 7 du terminal et les phrases plus complexes sont reconnues rapidement par le moteur de reconnaissance 14 du serveur. Le temps de traitement est d'autant plus optimisé que les VAD respectives présentent des caractéristiques adaptées, la VAD 5 du terminal étant par  
20 exemple conçue pour détecter rapidement des mots de commande et la VAD 12 du serveur étant conçue pour détecter rapidement des phrases.

Le contrôleur du terminal 8 détermine le signal à transmettre à destination du serveur par exemple en fonction de critères de contrôle. Ces critères peuvent par exemple être liés à l'application concernée, ou à des  
25 problématiques de charges des différents moyens de traitement au niveau du terminal et du serveur (les moyens de contrôle respectifs peuvent coopérer), ou encore à des problématiques de disponibilité de canaux de transmission voix ou de canaux de transmission données.

Par exemple, pour certains terminaux, le signal envoyé sera  
30 systématiquement un signal indiquant des paramètres de modélisation. Pour d'autres terminaux, le signal envoyé dépendra de l'application en cours.

Dans un mode de réalisation de l'invention, en cas de problèmes de disponibilité de canaux de données ou du module de calcul 6 du terminal

considéré, le contrôleur 8 est adapté pour transmettre le signal audio (d'origine ou après VAD). Ce signal audio pourra être transmis sur les canaux de transmission de signaux de voix disponibles.

5 Le façon de déterminer la forme finalement retenue et qui sera exploitée lors de la poursuite de l'application, entre une forme associée fournie par le module de reconnaissance du serveur et une forme associée fournie par celui du terminal peut s'effectuer sur la base de différents critères, qui peuvent varier d'un terminal à l'autre, mais aussi d'une application à l'autre ou d'un contexte donné à un autre.

10 Ces critères peuvent donner par exemple priorité à la reconnaissance effectuée au niveau du terminal, ou à la forme associée présentant le plus fort taux de probabilité, ou encore à la forme déterminée la plus rapidement. Ces critères de choix peuvent être intégrés par exemple dans les contrôleurs 8, 11 du terminal, ou du serveur.

15 Le serveur selon l'invention est également apte à effectuer de la reconnaissance de parole sur un signal transmis par un terminal ne disposant pas de moyens d'extraction de paramètres de modélisation, ni de moyens de reconnaissance (ou dont les moyens d'extraction ou de reconnaissance sont inactifs) et disposant éventuellement de VAD.

20 Dans un mode de réalisation, le moteur de reconnaissance 7 du terminal 2 est un programme exécutable téléchargé, par exemple depuis le serveur par des moyens classiques de transfert de données.

Avantageusement, pour une application donnée du terminal 2, des modèles de reconnaissance du terminal peuvent être téléchargés ou mis à jour  
25 au cours d'une session applicative connectée au réseau.

D'autres ressources logicielles utiles à la reconnaissance de parole peuvent aussi être téléchargés depuis le serveur 1, comme le module 6 de calcul de paramètres de modélisation ou le détecteur d'activité vocale 5.

30 D'autres exemples pourraient être décrits, mettant en œuvre par exemple des applications liées aux voitures, à l'électroménager, multimédia.

Comme présenté dans les exemples de réalisation ci-dessus décrits, un système selon l'invention permet d'utiliser dans un délai de traitement optimisé pour un résultat optimisé les différentes ressources nécessaires au

traitement de la reconnaissance de la parole et présentes au niveau du terminal (par exemple par téléchargement) et du serveur.

## REVENDICATIONS

1. Procédé de reconnaissance de parole distribuée, comportant au moins un terminal utilisateur et au moins un serveur aptes à communiquer entre eux par l'intermédiaire d'un réseau de télécommunications, selon lequel au niveau du terminal utilisateur, on réalise au moins les étapes suivantes :

- 5           - obtenir un signal audio à reconnaître ;  
          - calculer des paramètres de modélisation du signal audio à reconnaître; et  
          - tenter d'associer une forme mémorisée aux paramètres de modélisation; et
- 10          - indépendamment de l'étape de tentative d'association d'une forme mémorisée, émettre à destination du serveur un signal indiquant le signal audio à reconnaître;  
  
          et selon lequel au niveau du serveur, on réalise au moins les étapes suivantes :
- 15          - recevoir le signal émis par le terminal utilisateur ;  
          - tenter d'associer une forme mémorisée au signal reçu.

2. Procédé de reconnaissance de parole distribuée selon la revendication 1, selon lequel le signal émis par le terminal utilisateur à destination du serveur est sélectionné parmi au moins le signal audio à reconnaître et un signal indiquant les paramètres de modélisation ;

- 20           selon lequel si le signal reçu est de type audio, le serveur calcule des paramètres de modélisation du signal audio reçu, et tente d'associer une forme mémorisée aux paramètres de modélisation du signal audio reçu ;
- 25           et selon lequel si le signal reçu indique des paramètres de modélisation, le serveur tente d'associer une forme mémorisée auxdits paramètres de modélisation.

3. Procédé selon la revendication 1 ou la revendication 2, selon lequel l'obtention au terminal du signal à reconnaître comporte une détection d'activité vocale pour produire le signal audio à reconnaître sous forme de segments de parole extraits d'un signal audio d'origine en-dehors de périodes d'inactivité vocale.

4. Procédé selon la revendication 3, selon lequel le signal émis est un signal sélectionné parmi au moins le signal audio d'origine, le signal audio à reconnaître sous forme de segments extraits après détection vocale et le signal indiquant les paramètres de modélisation.

5. Procédé selon la revendication 2 à 4, selon lequel lorsque le signal reçu est de type audio :

- si le signal audio reçu est sous forme de segments de parole extraits après détection vocale, le serveur calcule des paramètres de modélisation du signal reçu et tente d'associer une forme mémorisée aux paramètres de modélisation du signal audio reçu ;

- sinon le serveur réalise une détection d'activité vocale appliquée au signal audio reçu pour produire un signal audio sous forme de segments de parole extraits du signal audio reçu en-dehors de périodes d'inactivité vocale, puis calcule des paramètres de modélisation du signal audio et tente d'associer une forme mémorisée aux paramètres de modélisation.

6. Procédé selon l'une quelconques des revendications précédentes, selon lequel on choisit la forme mémorisée associée déterminée au niveau du terminal quand elle existe.

7. Procédé selon les revendications 1 à 5 selon lequel on choisit la forme mémorisée associée déterminée le plus rapidement.

8. Procédé selon les revendications 1 à 5 selon lequel on choisit la forme mémorisée associée la meilleure selon un critère de choix défini.

5 9. Terminal utilisateur pour mettre en œuvre un procédé de reconnaissance de parole distribuée selon l'une des revendications 1 à 8, comportant :

- des moyens d'obtention d'un signal audio à reconnaître ;
- des moyens de calcul de paramètres de modélisation du signal audio; et
- 10 - des moyens de contrôle pour sélectionner un signal à émettre à destination du serveur parmi le signal audio à reconnaître et un signal indiquant les paramètres de modélisation calculés,
- des moyens de reconnaissance pour associer au moins une
- 15 forme mémorisée à des paramètres de modélisation calculés par les moyens de calcul.

10. Terminal utilisateur selon la revendication 9 dans lequel les moyens d'obtention du signal audio à reconnaître comprennent des moyens de détection d'activité vocale pour produire le signal à reconnaître sous forme de

20 segments de parole extraits d'un signal audio d'origine, en-dehors de périodes d'inactivité vocale.

11. Terminal utilisateur selon la revendication 10 dans lequel les moyens de contrôle sont adaptés pour sélectionner au moins un signal à

25 émettre à destination du serveur parmi le signal audio d'origine, le signal audio à reconnaître sous forme des segments de parole extraits par les moyens de détection d'activité vocale et le signal indiquant les paramètres de modélisation calculés.



12. Terminal utilisateur selon les revendications 9 à 11, dans lequel au moins une partie des moyens de calcul de paramètres et des moyens de reconnaissance est téléchargée depuis le serveur.

5           13. Terminal utilisateur selon les revendications 9 à 12, comportant des moyens de détermination de la forme mémorisée à choisir entre les formes mémorisées déterminées respectivement au terminal et au serveur.

10           14. Serveur pour mettre en œuvre un procédé de reconnaissance de parole distribuée selon l'une des revendications 1 à 8 comprenant :

- des moyens de réception en provenance d'un terminal utilisateur d'un signal sélectionné audit terminal ; et
- des moyens de reconnaissance pour associer au moins une forme mémorisée à des paramètres de modélisation en entrée.

15

15. Serveur selon la revendication 14 comprenant en outre :

- des moyens de calcul de paramètres de modélisation d'un signal d'entrée;
  - des moyens de contrôle pour commander les moyens de calcul et
- 20 les moyens de reconnaissance de façon à :

- lorsque le signal reçu par les moyens de réception est de type audio, activer les moyens de calcul de paramètres en leur adressant le signal sélectionné en tant que signal d'entrée, et adresser les paramètres calculés par les moyens de calcul aux moyens de reconnaissance en tant que paramètres

25 d'entrée, et,

- lorsque le signal sélectionné reçu par les moyens de réception indique des paramètres de modélisation, adresser lesdits paramètres indiqués aux moyens de reconnaissance en tant que paramètres d'entrée.

16. Serveur selon la revendication 15 comprenant en outre des moyens de détection d'activité pour produire le signal à reconnaître sous forme de segments de parole extraits d'un signal audio d'origine en-dehors de périodes d'inactivité vocale et dans lequel les moyens de contrôle sont adaptés  
5 pour commander les moyens de calcul de paramètres et les moyens de reconnaissance lorsque le signal reçu est de type audio de façon à :

- si le signal reçu de type audio est sous forme de segments de parole après détection d'activité vocale, activer les moyens de calcul de paramètres en leur adressant le signal reçu en tant que signal d'entrée, puis  
10 adresser les paramètres calculés par les moyens de calcul de paramètres aux moyens de reconnaissance en tant que paramètres d'entrée ;

- sinon activer les moyens de détection d'activité vocale du serveur en leur adressant le signal reçu en tant que signal d'entrée, puis adresser les segments extraits par les moyens de détection d'activité vocale aux moyens  
15 de calcul de paramètres en tant que paramètres d'entrée, puis adresser les paramètres calculés par les moyens de calcul de paramètres aux moyens de reconnaissance en tant que paramètres d'entrée ;

17. Serveur selon l'une des revendications 14 à 16, comprenant des  
20 moyens pour télécharger des ressources logicielles de reconnaissance vocale par l'intermédiaire du réseau de télécommunications à destination d'un terminal.

18. Serveur selon la revendication 17, dans lequel lesdites ressources  
25 comprennent au moins un module parmi : un module de VAD, un module de calcul de paramètres de modélisation d'un signal audio et un module de reconnaissance pour associer au moins une forme mémorisée à des paramètres de modélisation.

19. Serveur selon les revendications 14 à 18, comportant des moyens de détermination de la forme mémorisée à choisir entre les formes mémorisées déterminées respectivement au terminal et au serveur.

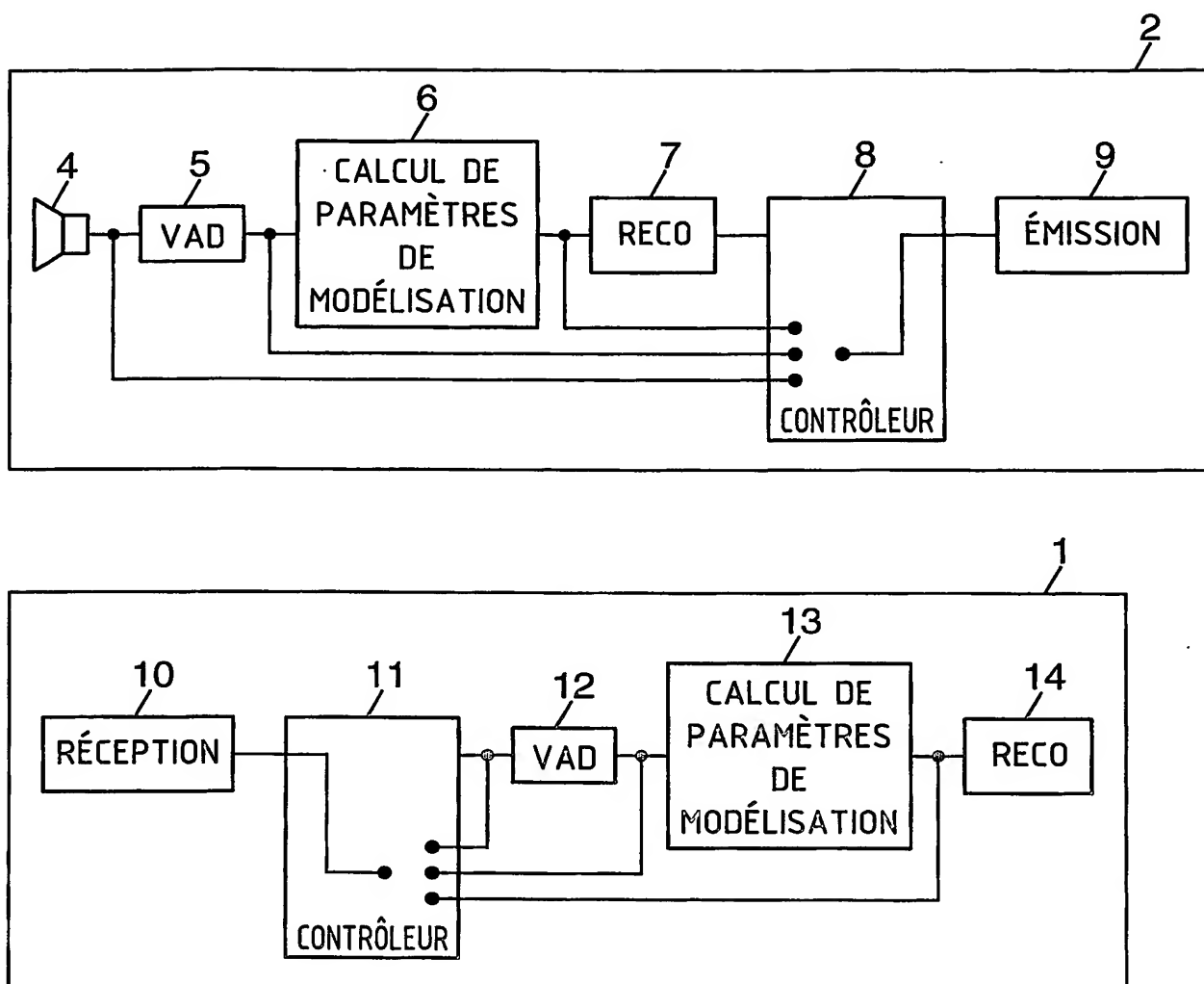


FIG. Unique

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/FR2004/000547

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 G10L15/28

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the International search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6 122 613 A (BAKER) 19 September 2000 (2000-09-19)	1, 2, 6-9, 12-15, 17-19
Y	abstract; figures 3-5, 10-12	3-5, 10, 11, 15, 16
Y	US 2003/004720 A1 (GARUDADRI ET AL) 2 January 2003 (2003-01-02) abstract; figure 3	3-5, 10, 11, 15, 16
A	US 6 487 534 B1 (THELEN ERIC ET AL) 26 November 2002 (2002-11-26) cited in the application the whole document	1-19

☐

Further documents are listed in the continuation of box C.

☒

Patent family members are listed in annex.

### \* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \* & \* document member of the same patent family

Date of the actual completion of the international search

28 July 2004

Date of mailing of the international search report

11/08/2004

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Quélavoine, R

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No  
PCT/FR2004/000547

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 6122613	A	19-09-2000	DE 69814589 D1	18-06-2003
			DE 69814589 T2	25-03-2004
			EP 0954848 A1	10-11-1999
			WO 9834217 A1	06-08-1998
US 2003004720	A1	02-01-2003	WO 02061727 A2	08-08-2002
US 6487534	B1	26-11-2002	AU 3165000 A	16-10-2000
			CN 1351745 T	29-05-2002
			WO 0058946 A1	05-10-2000
			WO 0058942 A2	05-10-2000
			EP 1181684 A1	27-02-2002
			EP 1088299 A2	04-04-2001
			JP 2002540477 T	26-11-2002
			JP 2002540479 T	26-11-2002

# RAPPORT DE RECHERCHE INTERNATIONALE

Demande Internationale No  
PCT/FR2004/000547

**A. CLASSEMENT DE L'OBJET DE LA DEMANDE**  
CIB 7 G10L15/28

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

**B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE**

Documentation minimale consultée (système de classification suivi des symboles de classement)  
CIB 7 G10L

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si réalisable, termes de recherche utilisés)  
EPO-Internal, WPI Data

**C. DOCUMENTS CONSIDERES COMME PERTINENTS**

Catégorie *	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	US 6 122 613 A (BAKER) 19 septembre 2000 (2000-09-19)	1,2,6-9, 12-15, 17-19
Y	abrégé; figures 3-5,10-12	3-5,10, 11,15,16
Y	----- US 2003/004720 A1 (GARUDADRI ET AL) 2 janvier 2003 (2003-01-02) abrégé; figure 3	3-5,10, 11,15,16
A	----- US 6 487 534 B1 (THELEN ERIC ET AL) 26 novembre 2002 (2002-11-26) cité dans la demande le document en entier	1-19

☐ Voir la suite du cadre C pour la fin de la liste des documents

☒ Les documents de familles de brevets sont indiqués en annexe

\* Catégories spéciales de documents cités:

- \*A\* document définissant l'état général de la technique, non considéré comme particulièrement pertinent
- \*E\* document antérieur, mais publié à la date de dépôt international ou après cette date
- \*L\* document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)
- \*O\* document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens
- \*P\* document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

- \*T\* document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention
- \*X\* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément
- \*Y\* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier
- \*&\* document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

28 juillet 2004

Date d'expédition du présent rapport de recherche internationale

11/08/2004

Nom et adresse postale de l'administration chargée de la recherche internationale  
Office Européen des Brevets, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Fonctionnaire autorisé

Quélavoine, R

# RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs

aux membres de familles de brevets

Demande Internationale No

PCT/FR2004/000547

Document brevet cité au rapport de recherche		Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
US 6122613	A	19-09-2000	DE 69814589 D1	18-06-2003
			DE 69814589 T2	25-03-2004
			EP 0954848 A1	10-11-1999
			WO 9834217 A1	06-08-1998
US 2003004720	A1	02-01-2003	WO 02061727 A2	08-08-2002
US 6487534	B1	26-11-2002	AU 3165000 A	16-10-2000
			CN 1351745 T	29-05-2002
			WO 0058946 A1	05-10-2000
			WO 0058942 A2	05-10-2000
			EP 1181684 A1	27-02-2002
			EP 1088299 A2	04-04-2001
			JP 2002540477 T	26-11-2002
			JP 2002540479 T	26-11-2002